We claim:

SUB A1

1. A computer-implemented method for quantifying gene relatedness for a plurality of candidate genes for which a plurality of gene expression level observations have been collected, the method comprising:

based on data comprising the plurality of gene expression level observations for the plurality of candidate genes, constructing a nonlinear model predicting gene expression;

predicting gene expression with the nonlinear model; and

measuring effectiveness of the nonlinear model in predicting gene expression, thereby quantifying gene relatedness for the plurality of candidate genes.

2. A computer-readable medium comprising computer-readable instructions for performing the method of claim 1.

3. The method of claim 1 wherein the nonlinear model accepts a plurality of predictive elements as inputs, wherein at least one of the predictive elements indicates whether a gene expression observation is associated with having applied a particular external stimulus to biological material.

4. The method of claim 1 wherein the nonlinear model accepts a plurality of predictive elements as inputs, wherein at least one of the predictive elements indicates whether a gene expression observation is associated with a particular cell state.

5. The method of claim 1 wherein the nonlinear model accepts a plurality of predictive elements as inputs, wherein at least one of the predictive elements indicates differential gene expression between two samples of biological material.

6. The method of claim 1 wherein the nonlinear model comprises a multivariate prediction function accepting two or more inputs to predict gene expression.

7. The method of claim 1 wherein measuring effectiveness of the nonlinear model comprises comparing observed gene expression to gene expression predicted by the nonlinear model.

8. The method of claim 1 wherein constructing the nonlinear model predicting gene expression comprises choosing a nonlinear model from a constrained set of nonlinear models.

9. The method of claim 1 wherein measuring the model's effectiveness comprises evaluating the model to estimate a coefficient of determination for an optimal model estimated by the model.

10. The method of claim 1 wherein the nonlinear model predicting gene expression is a full-logic model predicting gene expression for a predicted candidate gene, and the effectiveness of the model is measured by comparing predictions of gene expression for the predicted candidate gene by the model with observations of gene expression for the predicted candidate gene.

11. The method of claim 1 further comprising:

obtaining the data comprising the plurality of gene expression level observations from results of a plurality of cDNA microarray experiments measuring mRNA transcription levels for a plurality of genes in biological material.

12. The method of claim 1 wherein the data comprising a plurality of gene expression observations is divided into a training set of data and a test set of data, wherein

the nonlinear model predicting gene expression is generated via the training set data; and

effectiveness of the nonlinear model is measured via the test set of data.

13. The method of claim 12 wherein the training set of data is extended by randomly reordering and recycling gene expression observations.

14. The method of claim 1 wherein

a plurality of training data sets are repeatedly chosen from the data comprising a plurality of gene expression observations;

the nonlinear model is one of a plurality of models constructed from the plurality of training data sets; and

quantification of the relatedness for the plurality of candidate genes is measured by measuring average effectiveness of the plurality of models constructed from the plurality of training data sets.

15. The method of claim 1 further comprising:

to determine contribution of a predictive element to the quantification of relatedness, constructing an additional nonlinear model predicting gene expression, wherein the additional nonlinear model has a single input.

16. The method of claim 1 wherein the nonlinear model predicting gene expression is a truth table predicting a gene expression level for a predicted candidate gene from predictive elements comprising expression level observations for candidate genes other than the predicted candidate gene.

17. The method of claim 16 wherein the truth table comprises ternary discrete values.

18. The method of claim 16 wherein gene expression levels in the truth table are ternary discrete values.

19. The method of claim 16 wherein

the truth table comprises a plurality of rows for possible combinations of expression level observations for the candidate genes other than the predicted candidate gene; and

for at least one of the rows, the truth table indicates predicted gene expression for the predicted candidate gene with a thresholded weighted average of gene expression level observations associated with the row.

20. The method of claim 1 wherein the nonlinear model predicting gene expression is a neural network predicting gene expression.

21. The method of claim 20 wherein the neural network consists of one neuron which predicts a gene expression level for a single predicted candidate gene.

22. The method of claim 21 wherein the neuron is a ternary perceptron accepting predictive elements as inputs, wherein the predictive elements comprise gene expression levels indicated as one of three possible values: up, unchanged, and down.

23. The method of claim 22 further comprising:

displaying a three-dimensional graph representing the ternary perceptron with two planes separating points on the graph into points relating to like predicted values.

24. The method of claim 22 further comprising:

displaying a three-dimensional graph representing the ternary perceptron with objects at points in three-dimensional space within the graph, wherein axes of the graph relate to thresholded gene expression levels for three of the candidate genes.

25. The method of claim 24 wherein the objects are of a color indicating a predicted gene expression level.

26. The method of claim 24 wherein the objects are of a size indicating a number of observations related to a point on the graph.

27. The method of claim 1 wherein the data comprising a plurality of gene expression observations comprises gene expression level observations generated by subjecting sample biological material to an experimental condition and observing regulation of mRNA transcription levels for a plurality of genes in the biological material as a result of being subjected to the experimental condition.

28. The method of claim 27 wherein the data comprising a plurality of gene expression observations further comprises an indication of the experimental condition to which the biological material related to an observation was subjected and the indication is included in the model to predict gene expression.

29. A computer-implemented method for identifying genes related to a target gene by analyzing gene expression level observations for the genes, the method comprising:

based on the gene expression level observations, constructing multivariate nonlinear predictors that predict an expression level for the target gene, wherein the predictors accept gene expression levels for other genes as predictive elements;

estimating a coefficient of determination for sets of predictive elements and the target gene by comparing results of the multivariate nonlinear predictors with gene expression level observations for the target gene, wherein the predictive elements comprise expression level observations for genes other than the target gene; and

ranking the groups of genes other than the target gene by coefficient of determination to present the genes other than the target gene in order of likelihood of relatedness to the target gene.

30. The method of claim 29 further comprising:

indicating a proper subset of the genes having the highest likelihood of relatedness to the target gene.

31. A computer-implemented method for analyzing gene expression level observations for a set of genes comprising a target gene, the method comprising:

estimating a coefficient of determination for an optimal multivariate nonlinear model predicting gene expression of the target gene by constructing a multivariate nonlinear model from the gene expression level observations of gene expression for the target gene, wherein the optimal multivariate nonlinear model and the constructed multivariate nonlinear model predict gene expression of the target gene based on variables representing gene expression levels of genes other than the target gene.

32. The method of claim 31 wherein the optimal multivariate nonlinear model and the constructed multivariate nonlinear model predict gene expression based, at least in part, on inputs comprising an indication of a condition to which biological material relating to the observations has been subjected.

33. A method for identifying related genes out of a set of genes for which gene expression level observations have been collected, the method comprising:

for at least one predicted gene out of the set of genes, training an artificial intelligence function to predict gene expression for the predicted gene, wherein the artificial intelligence function takes one or more predictive elements as inputs and produces a gene expression level for the predicted gene as an output, wherein at least one of the predictive elements is a gene expression level for a gene other than the predicted gene; and

testing effectiveness of the artificial intelligence function in predicting expression of the predicted gene to rate relatedness of the predicted gene and at least one gene associated with the predictive elements.

34. The method of claim 33 wherein the artificial intelligence function takes a plurality of predictive elements as inputs.

35. The method of claim 33 wherein the predictive elements comprise a variable indicating biological material was subjected to an experimental condition.

36. For a plurality of observed genes for which expression levels have been observed, a method of presenting an analysis of the expression levels to assist in identifying related genes, the method comprising:

denoting a particular observed gene as a predicted gene;

for the predicted gene, constructing a plurality of nonlinear multivariate models predicting expression of the observed gene, wherein the nonlinear multivariate models comprise a variety of predictive elements chosen from permutations of expression levels of observed genes other than the predicted gene;

measuring effectiveness of the nonlinear multivariate models in predicting expression of the predicted gene to quantify relatedness between the predicted gene and the set of genes associated with the predictive elements of the models; and

presenting a quantification of relatedness between the predicted gene and a set of genes associated with the predictive elements of at least one of the models.

37. The method of claim 36 further comprising:

for a set of predictive elements and a predicted gene, displaying a graph indicating the amount of increase in the effectiveness of the model for each of the predictive elements.

38. The method of claim 36 wherein at least two of the plurality of nonlinear multivariate models predicting expression of the observed gene are implemented in specialized hardware circuits for predicting gene expression.

39. The method of claim 36 further comprising:

displaying a user interface for evaluating the analysis, wherein the user interface comprises display elements graphically indicating the relatedness of the predicted gene to a plurality of gene sets.

40. The method of claim 39 further comprising:

displaying only those display elements indicating sets of genes in which each gene in the set improves the relatedness.

41. The method of claim 39 further comprising:

accepting as input a set of one or more designated predictor genes;

accepting as input a threshold relatedness;

accepting as input a set of one or more designated predicted genes; and

limiting the display elements of the user interface to those sets of genes having as members the one or more designated predictor genes and having at least the threshold relatedness for the one or more designated predicted genes.

42. The method of claim 39 further comprising:

accepting as input a set of one or more designated predictor genes; and

limiting display elements of the user interface to those sets of genes having the one or more designated predictor genes.

43. The method of claim 42 further comprising:

accepting as input a threshold increase in relatedness; and

further limiting display elements of the user interface to those sets of genes for which addition of the one or more designated predictor genes increases the relatedness by at least the threshold increase in relatedness.

44. The method of claim 36 further comprising presenting a ranking of gene sets according to their relatedness, wherein the ranking indicates which genes are in the sets.

45. The method of claim 44 wherein the ranking further indicates contribution of individual predictive elements to the effectiveness of the models.

46. The method of claim 36 wherein the predictive elements comprise a variable indicative of whether biological material related to a gene expression observation has been subjected to a particular condition.

47. For a plurality of observed genes for which expression levels have been observed, a method of performing an analysis of the expression levels to assist in identifying related genes, the method comprising:

(a) for a plurality of the observed genes, denoting a particular observed gene as a predicted gene and performing at least (b) and (c);

(b) for the predicted gene, constructing a plurality of nonlinear multivariate models predicting expression of the predicted gene, wherein the nonlinear multivariate models have a variety of predictive elements chosen from permutations of expression levels of observed genes other than the predicted gene;

(c) measuring effectiveness of the nonlinear multivariate models in predicting expression of the predicted gene to provide a quantification of relatedness between the predicted gene and genes associated with the predictive elements of the models.

48. The method of claim 47 further comprising:

skipping designating genes having fewer than a defined number of changes in expression level as predicted genes.

49. The method of claim 47 further comprising:

displaying a user interface comprising display elements indicating gene relatedness for a plurality of genes associated with the predictive elements for a plurality of predicted genes.

50. A system for quantifying gene relatedness for a plurality of candidate genes for which a plurality of gene expression level observations have been collected, the system comprising:

SUB A2

means for constructing a nonlinear model predicting gene expression based on data comprising the plurality of gene expression level observations for the plurality of candidate genes;

means for predicting gene expression with the nonlinear model; and

means for measuring effectiveness of the nonlinear model in predicting gene expression, thereby quantifying gene relatedness for the plurality of candidate genes.

51. The method of claim 50 wherein the means for predicting gene expression is a specialized hardware circuit.

52. The method of claim 50 wherein the means for predicting gene expression is a decision tree.

53. The method of claim 50 wherein the means for predicting gene expression is a truth table chosen from a constrained set of truth tables.

54. A system for quantifying the relatedness of a set of genes, the system comprising:

SUB A3

a multivariate nonlinear predictor constructor operable to construct a multivariate nonlinear predictor based on gene expression level observations for a plurality of candidate genes; and

a multivariate nonlinear predictor tester operable to test the effectiveness of the multivariate nonlinear predictor to quantify relatedness for the plurality of candidate genes.

55. A computer user interface system for presenting results of a nonlinear multivariate prediction analysis of gene expression level data, the computer user interface system comprising:

a plurality of display elements, wherein each display element indicates a set of genes and a coefficient of determination of the set of genes for a target gene.

56. The computer user interface system of claim 55 wherein the plurality of display elements indicate a set of genes by displaying a different color for each gene in the set of genes.

57. The computer user interface system of claim 56 wherein the plurality of display elements indicate a set of genes by displaying a bar segment of a different color for each gene in the set of genes and the bar segment is of a size indicating contribution to the coefficient of determination by each gene in the set of genes.

58. The computer user interface system of claim 57 further comprising a contiguous display region for denoting a target gene associated with sets of genes, wherein the target gene is denoted with a color for the target gene and the sets of genes are grouped by target gene.

59. The computer user interface system of claim 55 wherein the user interface system accepts a threshold coefficient of determination to limit the display to sets of genes having at least the threshold coefficient of determination for a target gene.

60. The computer user interface system of claim 55 wherein the user interface accepts a threshold increase in coefficient of determination to limit the display to sets of genes having at least one gene whose inclusion in the set increases the coefficient of determination by at least the threshold increase in coefficient of determination.

61. A computer-implemented method of ranking the relatedness of a plurality of genes based on gene expression level observations associated with the plurality of genes, the method comprising:

based on the gene expression level observations, constructing a plurality of multivariate nonlinear predictors to predict the expression of a plurality of target genes out of the genes, wherein the multivariate nonlinear predictors comprise predictive elements comprising an observed gene, thereby associating the multivariate nonlinear predictor with the target gene and at least one observed gene;

testing effectiveness of the plurality of multivariate nonlinear predictors in predicting gene expression to quantify gene relatedness between the genes associated with the predictors by estimating a coefficient of determination; and

displaying a ranked list of gene relatedness among the genes as determined by testing the plurality of multivariate nonlinear predictors.

ADD A47 add B1